

CLMPTO

6/19/02

TB

CLAIMS 1-17 CANCELLED

18. (NEW) A method for conditioning a database of words formed of graphemes and phonemes, for automatic speech processing using the database, said method comprising:  
    automatically assigning the graphemes to the phonemes to form  
grapheme/phoneme pairs for the words in the database having an identical number of  
graphemes and phonemes;

automatically assigning the graphemes to the phonemes for each word in the database having more graphemes than phonemes, to form grapheme/phoneme pairs until an assignment error is detected based on previously determined assignments, or no phonemes remain for the word, then combining a plurality of graphemes in the word to form a grapheme unit and assigning the grapheme unit to one of the phonemes;

automatically assigning the graphemes to the phonemes for each word in the database having fewer graphemes than phonemes, by assigning a single grapheme in the word to a plurality of phonemes combined to form a phoneme unit, with remaining graphemes and phonemes in the word corresponding to previously assigned grapheme/phoneme pairs;

automatically reexamining remaining words for previously determined grapheme/phoneme pairs and at least one of phoneme units and grapheme units determined previously; and

automatically erasing, at least after said assigning for the words in the database having an identical number of graphemes and phonemes, any assignment which contradicts another assignment determined to be valid.

19. (NEW) The method as claimed in claim 18, wherein said removing is performed after said assigning is performed for all the words in the database having more graphemes than phonemes and after said assigning is performed for all the words in the database having fewer graphemes than phonemes.

20. (NEW) The method as claimed in claim 19, further comprising determining assignments to be valid for each grapheme/phoneme pair having a frequency of occurrence above a predetermined threshold value.

21. (NEW) The method as claimed in claim 20, further comprising removing each word in a list of completed words, containing an assignment subjected to said erasing.

22. (NEW) The method as claimed in claim 21, wherein said assigning for words having fewer graphemes than phonemes uses phoneme units that do not combine a vowel with a consonant.

23. (NEW) The method as claimed in claim 21, wherein said assigning for words having fewer graphemes than phonemes first uses phoneme units that do not combine a vowel

with a consonant and, if any words having fewer graphemes than phonemes remain, then uses phoneme units combining vowels with consonants.

24. (NEW) The method as claimed in claim 21, wherein said assigning of the graphemes to the phonemes for words having more graphemes than phonemes comprises:

assigning graphemes and phonemes to form the grapheme/phoneme pairs, starting from one end of each word until a last assigned grapheme is assigned to a final phoneme;

combining each unassigned grapheme with the last assigned grapheme to form an ending grapheme unit; and

assigning the ending grapheme unit to the final phoneme of the word.

25. (NEW) The method as claimed in claim 24, wherein said assigning of the graphemes to the phonemes for words having more graphemes than phonemes further comprises:

assigning the graphemes and the phonemes to form the grapheme/phoneme pairs, from a first end of each word towards a second end of the word, until a first assignment contradiction is detected based on previous assignments determined to be valid;

combining a last grapheme in a longest chain of graphemes free of assignment contradictions with a first unassigned grapheme to form an intermediate grapheme unit;

assigning the intermediate grapheme unit to a last assigned phoneme in the longest chain; and

continuing said assigning of the graphemes and the phonemes to form the grapheme/phoneme pairs while checking for another assignment contradiction and, each time a next unassigned grapheme, immediately following the intermediate grapheme unit, produces another assignment contradiction, adding the next unassigned grapheme to the intermediate grapheme unit and reassigning the intermediate grapheme unit to the last assigned phoneme in the longest chain.

26. (NEW) The method as claimed in claim 25, further comprising for each of the words in which said continuing is unable to resolve all assignment contradictions:

using said assigning of the graphemes and the phonemes to form the grapheme/phoneme pairs from the first end of the word to the first assignment contradiction;

assigning the graphemes and the phonemes to form the grapheme/phoneme pairs, from the second end of the word towards the first end of the word, until a single unassigned phoneme remains;

combining all unassigned graphemes in the word to form a single grapheme unit;  
and

assigning the single grapheme unit to the single unassigned phoneme.

27. (NEW) A method for training a neural network to assign graphemes to phonemes for automatic speech processing, comprising:

conditioning a database of words formed of graphemes and associated phonemes by

assigning the graphemes to the phonemes to form grapheme/phoneme pairs for the words in the database having an identical number of graphemes and phonemes;

assigning the graphemes to the phonemes for each word in the database having more graphemes than phonemes, to form grapheme/phoneme pairs until an assignment error is detected based on previously determined assignments, or no phonemes remain for the word, then combining a plurality of graphemes in the word to form a grapheme unit and assigning the grapheme unit to one of the phonemes;

assigning the graphemes to the phonemes for each word in the database having fewer graphemes than phonemes, by assigning a single grapheme in the word to a plurality of phonemes combined to form a phoneme unit, with remaining graphemes and phonemes in the word corresponding to previously assigned grapheme/phoneme pairs;

reexamining remaining words for previously determined grapheme/phoneme pairs and at least one of phoneme units and grapheme units determined previously; and

erasing, at least after said assigning for the words in the database having an identical number of graphemes and phonemes, any assignment which contradicts another assignment determined to be valid; and

training the neural network by inputting the graphemes at input nodes and the associated phonemes at an output node of the neural network for at least some of the words in the database.

28. (NEW) The method as claimed in claim 27, wherein said removing is performed after said assigning is performed for all the words in the database having more graphemes than

phonemes and after said assigning is performed for all the words in the database having fewer graphemes than phonemes.

29. (NEW) The method as claimed in claim 28, further comprising determining assignments to be valid for each grapheme/phoneme pair having a frequency of occurrence above a predetermined threshold value.

30. (NEW) The method as claimed in claim 29, further comprising removing each word in a list of completed words, containing an assignment subjected to said erasing.

31. (NEW) The method as claimed in claim 30, wherein said assigning for words having fewer graphemes than phonemes uses phoneme units that do not combine a vowel with a consonant.

32. (NEW) The method as claimed in claim 30, wherein said assigning for words having fewer graphemes than phonemes first uses phoneme units that do not combine a vowel with a consonant and, if any words having fewer graphemes than phonemes remain, then uses phoneme units combining vowels with consonants.

33. (NEW) The method as claimed in claim 30, wherein said assigning of the graphemes to the phonemes for words having more graphemes than phonemes comprises:  
assigning graphemes and phonemes to form the grapheme/phoneme pairs,  
starting from one end of each word until a last assigned grapheme is assigned to a final phoneme;

combining each unassigned grapheme with the last assigned grapheme to form an ending grapheme unit; and

assigning the ending grapheme unit to the final phoneme of the word.

34. (NEW) The method as claimed in claim 33, wherein said assigning of the graphemes to the phonemes for words having more graphemes than phonemes further comprises:

assigning the graphemes and the phonemes to form the grapheme/phoneme pairs, from a first end of each word towards a second end of the word, until a first assignment contradiction is detected based on previous assignments determined to be valid;

combining a last grapheme in a longest chain of graphemes free of assignment contradictions with a first unassigned grapheme to form an intermediate grapheme unit;  
assigning the intermediate grapheme unit to a last assigned phoneme in the longest chain; and

continuing said assigning of the graphemes and the phonemes to form the grapheme/phoneme pairs while checking for another assignment contradiction and, each time a next unassigned grapheme, immediately following the intermediate grapheme unit, produces another assignment contradiction, adding the next unassigned grapheme to the intermediate grapheme unit and reassigning the intermediate grapheme unit to the last assigned phoneme in the longest chain.

35. (NEW) The method as claimed in claim 34, further comprising for each of the words in which said continuing is unable to resolve all assignment contradictions:

using said assigning of the graphemes and the phonemes to form the grapheme/phoneme pairs from the first end of the word to the first assignment contradiction;

assigning the graphemes and the phonemes to form the grapheme/phoneme pairs, from the second end of the word towards the first end of the word, until a single unassigned phoneme remains;

combining all unassigned graphemes in the word to form a single grapheme unit;  
and

assigning the single grapheme unit to the single unassigned phoneme.

36. (NEW) A method for assigning graphemes to phonemes for synthesizing speech, comprising:

conditioning a database of words formed of graphemes and associated phonemes by

assigning the graphemes to the phonemes to form grapheme/phoneme pairs for the words in the database having an identical number of graphemes and phonemes;

assigning the graphemes to the phonemes for each word in the database having more graphemes than phonemes, to form grapheme/phoneme pairs until an assignment error is detected based on previously determined assignments, or no phonemes remain for the word, then combining a plurality of graphemes in the word to form a grapheme unit and assigning the grapheme unit to one of the phonemes;

assigning the graphemes to the phonemes for each word in the database having fewer graphemes than phonemes, by assigning a single grapheme in the word to a plurality of phonemes combined to form a phoneme unit, with remaining graphemes and phonemes in the word corresponding to previously assigned grapheme/phoneme pairs;

reexamining remaining words for previously determined grapheme/phoneme pairs and at least one of phoneme units and grapheme units determined previously;

erasing, at least after said assigning for the words in the database having an identical number of graphemes and phonemes, any assignment which contradicts another assignment determined to be valid; and

training a neural network by inputting the graphemes at input nodes and the associated phonemes at an output node of the neural network for at least some of the words in the database;

inputting, for each word to be synthesized, an input pattern formed of a letter to be assigned and, if present, at least one letter preceding the letter to be assigned and one subsequent letter in the word; and

generating, from the neural network, an output pattern of one phoneme corresponding to the letter to be assigned.

37. (NEW) The method as claimed in claim 36, wherein said inputting includes, when possible, a plurality of letters preceding and subsequent to the letter to be assigned.

38. (NEW) The method as claimed in claim 37, wherein said inputting includes, when possible, three preceding letters and three subsequent letters.

39. (NEW) The method as claimed in claim 38, wherein the input pattern includes, when possible, an immediately previous output pattern.

40. (NEW) The method as claimed in claim 39, wherein the output pattern indicates a number of the letters combined to form the grapheme unit in the input pattern.

41. (NEW) A method for assigning phonemes to graphemes for recognition of speech, comprising:

conditioning a database of words formed of graphemes and associated phonemes by

assigning the graphemes to the phonemes to form grapheme/phoneme pairs for the words in the database having an identical number of graphemes and phonemes;

assigning the graphemes to the phonemes for each word in the database having more graphemes than phonemes, to form grapheme/phoneme pairs until an assignment error is detected based on previously determined assignments, or no phonemes remain for the word, then combining a plurality of graphemes in the word to form a grapheme unit and assigning the grapheme unit to one of the phonemes;

assigning the graphemes to the phonemes for each word in the database having fewer graphemes than phonemes, by assigning a single grapheme in the word to a plurality of phonemes combined to form a phoneme unit, with remaining graphemes and phonemes in the word corresponding to previously assigned grapheme/phoneme pairs;

reexamining remaining words for previously determined grapheme/phoneme pairs and at least one of phoneme units and grapheme units determined previously;

erasing, at least after said assigning for the words in the database having an identical number of graphemes and phonemes, any assignment which contradicts another assignment determined to be valid; and

training a neural network by inputting the phonemes at input nodes and the associated graphemes at an output node of the neural network for at least some of the words in the database;

inputting, to the neural network, an input pattern formed of at least a phoneme to be assigned and, if present, a preceding phoneme and a subsequent phoneme; and

generating, from the neural network, an output pattern having one letter corresponding to the phoneme to be assigned.

42. (NEW) A neural network for automatic speech processing trained by conditioning a database of words formed of graphemes and associated phonemes by

assigning the graphemes to the phonemes to form grapheme/phoneme pairs for the words in the database having an identical number of graphemes and phonemes;

assigning the graphemes to the phonemes for each word in the database having more graphemes than phonemes, to form grapheme/phoneme pairs until an assignment



error is detected based on previously determined assignments, or no phonemes remain for the word, then combining a plurality of graphemes in the word to form a grapheme unit and assigning the grapheme unit to one of the phonemes;

assigning the graphemes to the phonemes for each word in the database having fewer graphemes than phonemes, by assigning a single grapheme in the word to a plurality of phonemes combined to form a phoneme unit, with remaining graphemes and phonemes in the word corresponding to previously assigned grapheme/phoneme pairs;

reexamining remaining words for previously determined grapheme/phoneme pairs and at least one of phoneme units and grapheme units determined previously; and

erasing, at least after said assigning for the words in the database having an identical number of graphemes and phonemes, any assignment which contradicts another assignment determined to be valid; and

training the neural network by inputting the graphemes at input nodes and the associated phonemes at an output node of the neural network for at least some of the words in the database.

43. (NEW) An electronically readable data medium storing at least one computer program to control a processor to perform a method for conditioning a database of words formed of graphemes and phonemes, for automatic speech processing using the database, said method comprising:

automatically assigning the graphemes to the phonemes to form grapheme/phoneme pairs for the words in the database having an identical number of graphemes and phonemes;

automatically assigning the graphemes to the phonemes for each word in the database having more graphemes than phonemes, to form grapheme/phoneme pairs until an assignment error is detected based on previously determined assignments, or no phonemes remain for the word, then combining a plurality of graphemes in the word to form a grapheme unit and assigning the grapheme unit to one of the phonemes;

automatically assigning the graphemes to the phonemes for each word in the database having fewer graphemes than phonemes, by assigning a single grapheme in the word to a plurality of phonemes combined to form a phoneme unit, with remaining graphemes and phonemes in the word corresponding to previously assigned grapheme/phoneme pairs;

automatically reexamining remaining words for previously determined grapheme/phoneme pairs and at least one of phoneme units and grapheme units determined previously; and

automatically erasing, at least after said assigning for the words in the database having an identical number of graphemes and phonemes, any assignment which contradicts another assignment determined to be valid.